**Homework #8 (Due Oct 22 11:59 PM)**

IST 3420 - Fall 2017, Chen

**Name**: \_\_\_\_\_\_\_\_\_Adam Forestier\_\_\_\_\_\_\_\_\_\_\_

**Plot MST Out-of-State Enrollment on US Map (20 points)**

**Instruction:**

Finish the following tasks to plot MST out-of-state enrollment data on a US map. The following plot shows an example of the required plot. Your final plot can be different from the sample plot.



***Tasks:***

Download the MST enrollment file “mst\_enrollment2015.csv”.

Read the data file into R.

Use dplyr::filter() to remove all non-US data. (1 point)

Use dplyr::filter() to remove data of Missouri (in-state enrollment). (1 point)

Use tidyr package to reshape the dataset by transforming it from wide format to long format, using “Year” as key and “Enroll” as value. (2 points)

Use dplyr package to aggregate enrollment data into state level, i.e. calculate total enrollment for each state. (2 points)

List top 6 states from which the out-of-state students come (year 2011 through 2015). Fill the following table. (2 points)

|  |  |  |
| --- | --- | --- |
| Rank | State (Non-Missouri) | Total Enrollment at MST |
| 1 | Illinois | 2532 |
| 2 | Texas | 697 |
| 3 | Kansas | 621 |
| 4 | California | 276 |
| 5 | Washington | 270 |
| 6 | Oklahoma | 267 |

Use ggplot2 package to draw the MST out-of-state enrollment plot on the US states map. Paste your plot in the following box. (10 points)

|  |
| --- |
|  |

Paste your R markdown code in the following box (1 point).

|  |
| --- |
| ---  title: "Homework 8"  author: "Adam Forestier"  date: "October 22, 2017"  output: html\_document  ---  ```{r setup, include=FALSE}  knitr::opts\_chunk$set(echo = TRUE, message = F)  ```  Clean Environment  ```{r}  rm(list = ls())  ```  Load Packages  ```{r}  require(dplyr)  require(magrittr)  require(ggplot2)  require(tidyr)  require(maps)  ```  # Read MST Enrollment data  ```{r}  mst <- read.csv("mst\_enrollment2015.csv")  # Show structure  str(mst)  ```  # Remove not US Data  ```{r}  mst\_us <- mst %>% dplyr::filter(mst$Country == "United States")  ```  # Remove Missouri Data  ```{r}  mst\_us <- mst\_us %>% dplyr::filter(mst\_us$State != "Missouri")  ```  Load State Map Data  ```{r}  # Create a data frame of map data for all states  us\_states <- map\_data("state")  # List all states in map dataset  map\_state <- unique(us\_states$region)  map\_state  ```  We see that Armed Forces Europe, Armed Forces Pacific, Hawaii, and Alaska aren't in the map package so we need to remove them  ```{r}  mst\_us <- mst\_us %>% dplyr::filter(mst\_us$State != "Armed Forces Europe")  mst\_us <- mst\_us %>% dplyr::filter(mst\_us$State != "Armed Forces Pacific")  mst\_us <- mst\_us %>% dplyr::filter(mst\_us$State != "Alaska")  mst\_us <- mst\_us %>% dplyr::filter(mst\_us$State != "Hawaii")  ```  # Reshape the dataset by transforming it from wide format to long format  ```{r}  mst\_us\_long <- mst\_us %>% gather(Year, Enroll, X2011, X2012, X2013, X2014, X2015)  ```  lower capitilization on State names  ```{r}  mst\_us\_long$State <- tolower(mst\_us\_long$State)  ```  # Aggregate enrollment data into state level, i.e. calculate total enrollment for each state using dplyr  ```{r}  # Aggregate data  by\_state <- group\_by(mst\_us\_long, State)  enrollment\_count <- summarise(by\_state, pop = sum(Enroll))  # Show top 6 enrollment states  head(enrollment\_count[order(enrollment\_count$pop, decreasing = T),])  ```  # MST out-of-state enrollment plot on the US states map  Merge Data Sets  ```{r}  # Rename Columns of us\_states for merging purposes  colnames(us\_states) <- c("long", "lat", "group", "order", "State", "subregion")  # Merge Enrollment\_Count and us\_states for map  pop\_map <- merge(us\_states, enrollment\_count, by="State")  ```  Create Map Visualization  ```{r}  ggplot() +  geom\_polygon(data = pop\_map,  aes(x=long, y=lat, group = group, fill=pop\_map$pop),  colour="black") +  scale\_fill\_continuous(low = "darkslategray1", high = "darkblue", guide="colorbar") +  theme\_bw() +  labs(fill = "Enrollment",  title = "Missouri S&T Enrollment by State",  x="", y="") +  scale\_y\_continuous(breaks=c()) +  scale\_x\_continuous(breaks=c()) +  theme(panel.border = element\_blank())  ``` |

Upload this document with your answers to “Homework 8” on Canvas.

Upload your R Markdown file to “Homework 8” on Canvas. (1 point)